

NCBI Molecular Biology Resources

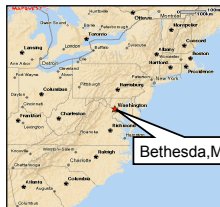
Part 1: NCBI Databases and Entrez

December 2009

NCBI Databases and Entrez

- About NCBI
- Molecular Databases
- The Entrez system and Discovery
- Using Entrez relationships
- Example search: Finding sequences, genomic information and structures for human MLH1 gene and products

The National Center for Biotechnology Information



Created in 1988 as a part of the National Library of Medicine at NIH

- Establish public databases
- Research in computational biology
- Develop software tools for sequence analysis
- Disseminate biomedical information

Web Access: www.ncbi.nlm.nih.gov

Resources

- NCBI Home
- All Resources (A-Z)
- Literature
- DNA & RNA
- Proteins
- Sequence Analysis
- Genes & Expression
- Genomes
- Maps & Markers
- Domains & Structures
- Genetics & Medicine
- Taxonomy
- Data & Software
- Training & Tutorials
- Homology
- Small Molecules
- Variation

GETTING STARTED

- Site Map
- NCBI Help Manual
- NCBI Handbook
- Training & Tutorials

RESOURCES

- Literature
- DNA & RNA
- Proteins
- Sequence Analysis
- Genes & Expression
- Genomes
- Maps & Markers
- Domains & Structures
- Genetics & Medicine
- Taxonomy
- Data & Software
- Training & Tutorials
- Homology
- Small Molecules
- Variation

POPULAR

- PubMed
- PubMed Central
- Bookshelf
- BLAST
- Gene
- Nucleotide
- Protein
- GenBank
- RefSeq
- Map Viewer
- Human Genome
- Mouse Genome
- Influenza Virus
- Primer-BLAST
- Short Read Archive

FEATURED

- GenBank
- RefSeq
- Map Viewer
- Human Genome
- Mouse Genome
- Influenza Virus
- Primer-BLAST
- Short Read Archive

NCBI INFORMATION

- About NCBI
- Research at NCBI
- NCBI Newsletter
- NCBI FTP Site
- Contact Us

Common footer

NCBI Databases and Services

- GenBank primary sequence database
- Free public access to biomedical literature
 - PubMed free Medline (3 million searches per day)
 - PubMed Central full text online access
- Entrez integrated molecular and literature databases
- BLAST highest volume sequence search service (100 – 200 K searches per day)
- VAST structure similarity searches
- Software and databases for download

NCBI Field Guide

Types of Molecular Databases

- Primary Databases
 - Original submissions by experimentalists
 - Content controlled by the submitter
 - Examples: GenBank, Trace, SRA, SNP, GEO
- Derivative Databases
 - Built from primary data
 - Content controlled by third party (NCBI)
 - Examples: NCBI Protein, Refseq, TPA, RefSNP, GEO datasets, UniGene, Homologene, Structure, Conserved Domain

NCBI Field Guide

Sequence Databases

NCBI Field Guide

Sequence Databases at NCBI

- Primary
 - GenBank: NCBI's primary sequence database
 - Trace Archive: reads from capillary sequencers
 - Sequence Read Archive: next generation data
- Derivative
 - GenPept (GenBank translations)
 - Outside Protein (UniProt—Swiss-Prot, PDB)
 - NCBI Reference Sequences (RefSeq)

NCBI Field Guide

NCBI Field Guide

- NCBI Field Guide

NCBI Field Guide



NCBI Field Guide

NCBI Field Guide

NCBI Field Guide

- NCBI Field Guide

- NCBI Field Guide

Expressed Sequence Tags in Entrez

Items 1 - 20 of 63441036 (Page) 1 of 3172052 Next

Top Organisms [Tree]

Item	Sequence Description	Length	Organism
1	AZ061310D Agropyron mongolicum Keng Transcript-derived fragment Agropyron mongolicum cDNA, mRNA sequence	148 bp linear mRNA	Homo sapiens (8296280)
2	WSR443 Withania somnifera cDNA, mRNA sequence	365 bp linear mRNA	Mus musculus (4521144)
3	WSR408 Withania somnifera cDNA, mRNA sequence	247 bp linear mRNA	Zen mays (2318798)
4	WSR428 Withania somnifera cDNA, mRNA sequence	280 bp linear mRNA	Bos taurus (552571)

Organism	Records
Total	63 million records
Human	8.3 million
Mouse	4.9 million
Maize	2.0 million
Cow	1.6 million
Pig	1.5 million
Arabidopsis	1.5 million
Zebrafish	1.5 million
Soybean	1.4 million
Xenopus tropicalis	1.3 million
Rice (all)	1.2 million
Ciona intestinalis	1.2 million
Wheat	1.0 million
Rat	1.0 million

Whole Genome Shotgun Projects

ftp.ncbi.nih.gov/genbank/wgs/

- >900 Projects
- >800 Taxa
 - 585 Bacteria
 - 8 Archaea
 - 17 metagenomes
 - 255 eukaryotes
 - 86 fungi
 - 89 animals
 - 7 flowering plants

Derivative Sequence Databases

GenPept: GenBank CDS translations

```

FEATURES             Location/Qualifiers
     source            1..2484
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /chromosome="3"
                        /map="3p22-p23"
     gene              1..2484
                        /gene="DNA mismatch repair protein homolog"
     CDS                22..2292
                        /gene="DNA mismatch repair protein homolog"
                        /note="homolog of S. cerevisiae PMS1 (Swiss-Prot Accession
                        Number P14242), S. cerevisiae MLH1 (GenBank Accession
                        Number U07187), E. coli MUTL (Swiss-Prot Accession Number
                        P23367), Salmonella typhimurium MUTL (Swiss-Prot Accession
                        Number P14161), Streptococcus pneumoniae (Swiss-Prot
                        Accession Number P14160)"
                        /codon_start=1
                        /product="DNA mismatch repair protein homolog"
                        /protein_id="AA050285.1"
                        /db_xref="GI:463989"
                        /translation="MSFVAGVIRRLDETVDVNNIAAGEVIQRPANAIEKMIENCLDAKS
                        TSIIQIVKEGGLKLIQDNGTGIRKEDLDIVCFRTTSKLQSFEDLASISTYGFGRGE
                        ALASISHVAHVTTITKTADGKCAIRASYSDDGLKAPPKPCAGNQTGITVDLFYNIA
                        TRRKALNPFSEYGVKILEVGVGRVSHNAGISFVKKQGETVADVTRTPNASTVDNIRS"
    
```


NCBI Field Guide

>gi|5542073|pdb|1B63|A Chain A, Mutl Complexed With Adnp
 SMPALPVLQPLQANIAAGEVFRSPVKELVNSDAGATIRIDDERGAALIRIDNGCGKKDEL
 ALAARHATSKIASIDLEALISLPGFGALASISVSRLT'LSRTAEQQAQYAEGRMNVTPKPA
 HPVGT'LEVLDFYNT'PARKFLRTEKTEFNHIDEIRRIARALFDVTINLSHGKIVRQYRAVPEGGQK
 ERLGAGICGTAFLEAGLEAWEQBGDILRGLWADPNHT'PALAEIQ'CYCVLSGRMRDRILINHAIRQACED
 KLGADGQPAFVLELIDHPQVNDVHPKAEVHGRSRLVDEIT'FQGVLSVLQ

NCBI Field Guide

- <ftp://ftp.ncbi.nih.gov/refseq/release/>

NCBI Field Guide

NM_123456	Curated mRNA
NP_123456	Curated Protein
NR_123456	Curated non-coding RNA
XM_123456	Predicted mRNA
XP_123456	Predicted Protein
XR_123456	Predicted non-coding RNA
<u>Gene Records</u>	
NG_123456	Reference Genomic Sequence
<u>Chromosome</u>	
NC_123455	Microbial replicons, organelle genomes, human chromosomes
	Alternate assemblies
AC_123455	
<u>Assemblies</u>	
NT_123456	Contig
NW_123456	WGS Supercontig

NCBI Field Guide

<ol style="list-style-type: none"> Human apolipoprotein E (epsilon-4 allele) gene, complete cds M10265.1 GI:178852 Human mRNA fragment for apolipoprotein E (apo E) 528 bp linear mRNA X00159.1 GI:28808 H sapiens mRNA 275 linear nt Z70760.1 GI:1263 Homo sapiens 1,023 bp linear AK034896.1 GI:19 Human apolipoprotein E mRNA, complete cds 1,157 bp linear mRNA M12829.1 GI:178848 Homo sapiens preapolipoprotein E (APOE) mRNA, complete cds 1,156 bp linear mRNA K03299.1 GI:178850 Homo sapiens apolipoprotein E, mRNA (cDNA clone MGC:1571 IMAGE:3355712), complete cds 1,186 bp linear mRNA U00004.1 GI:178851 	<div> <div> Homo sapiens apolipoprotein E (APOE), mRNA </div> <div> 1,223 bp linear mRNA </div> <div> NM_000041.2 GI:48762938 </div> </div>
--	---

Expressed Sequences

UniGene
GEO

NCBI Expressed Sequences

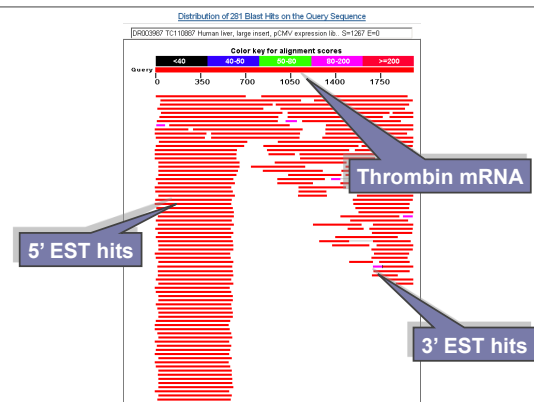
67,920,384 mRNA sequences
65,906,124 GenBank
(63,832,762 EST Division)
2,012,137 Reference Sequences

What is UniGene?

A gene-oriented view of sequence entries

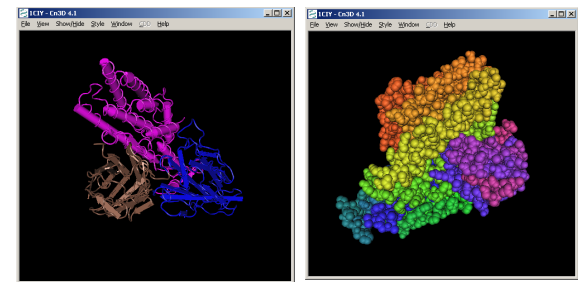
- MegaBlast based automated sequence clustering
- Now informed by genome hits
- Nonredundant set of gene oriented clusters
- Each cluster a unique gene
- Information on tissue types and map locations
- Includes known genes and uncharacterized ESTs
- Useful for gene discovery and selection of mapping reagents

EST hits: Human mRNA



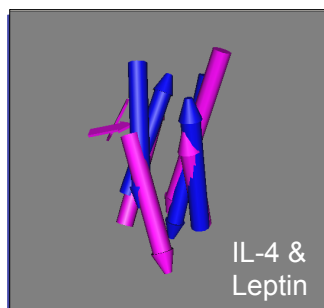
NCBI Field Guide

- ### Cn3D 4.1: *Bacillus thuringiensis* Toxin

[illegible]

NCBI Field Guide

align the vectors



NCBI Field Guide

- 9

NCBI Field Guide

- NCBI Field Guide

NCBI Field Guide



NCBI Field Guide



NCBI Field Guide

- NCBI Field Guide

NCBI Field Guide

- NCBI Field Guide

NCBI Field Guide

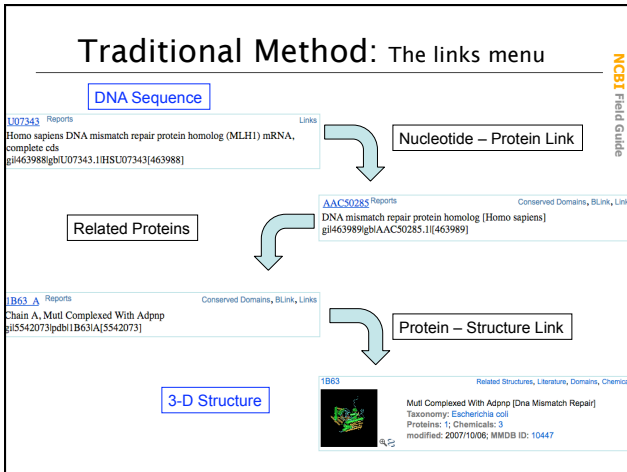
November 2009

NCBI Field Guide

NCBI Field Guide

NCBI Field Guide





The Problem

- Rapidly growing databases with complex and changing relationships
- Rapidly changing interfaces to match the above

Result

- Many people don't know:
 - Where to begin
 - Where to click on a Web page
 - Why it might be useful to click there

Goals of the Discovery Initiative

- Easier to use interfaces
- Promote higher quality resources
 - Gene
 - RefSeqs
- Expose the power and utility of pre-computed similarities and pre-compiled links

Discovery Components in Entrez

- Database Ads – direct to related information in other database
- Sensors – point to other databases or special search tools where the query is more relevant
- Analysis tools – access to live analysis results

Database Searching with Entrez

- ◆ Using limits and field restriction to find human MutL homolog
- ◆ Linking and neighboring with MutL
- ◆ Mapping SNPs onto structure

Global NCBI (Entrez) Search

Global Entrez Search Results

Nucleotide Sequences

Nucleotide database in three parts

- EST: expressed sequence tags
- GSS: genome survey sequences
- Nucleotide: everything else

Core Nucleotide Results with Gene Preview

Found 5104813 nucleotide sequences. Nucleotide [941100] EST [1352867] GSS [310846]

Display Summary Show 20 Sort By Send to

All: 941100 Bacteria: 25600 RefSeq: 50115 mRNA: 240717

This search in Gene shows 700 results, including:

- PTPRJ** (Homo sapiens): protein tyrosine phosphatase, receptor type, J
- MSH2** (Homo sapiens): mutS homolog 2, colon cancer, nonpolyposis type 1 (E. coli)
- MLH1** (Homo sapiens): mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)

Items 1 - 20 of 941100

Gene Preview
More relevant results

Taxonomy Filters

Top Organisms [Tree]

- Homo sapiens (482635)
- synthetic construct (118957)
- Mus musculus (69474)
- Danio rerio (29495)
- Xenopus laevis (15332)
- Rattus norvegicus (14639)
- Helicobacter pylori (13098)
- Xenopus (Silurana) tropicalis (11721)
- Human immunodeficiency virus 1 (11699)
- uncultured bacterium (9877)
- Helicobacter pylori
- HPKX_438_CAMC1 (7534)
- HPKX_438_AGOC1 (5206)
- Grossmannia clavigena kv1407 (4996)
- Pongo abelii (4643)
- Drosophila melanogaster (1614)
- Felis catus (1176)
- Canis lupus familiaris (1135)
- All other taxa (24999)

Advanced Search Options

Found 5104813 nucleotide sequences. Nucleotide [941100] EST [1352867] GSS [310846]

Display Summary Show 20 Sort By Send to

All: 941100 Bacteria: 25600 RefSeq: 50115 mRNA: 240717

This search in Gene shows 700 results, including:

- PTPRJ** (Homo sapiens): protein tyrosine phosphatase, receptor type, J
- MSH2** (Homo sapiens): mutS homolog 2, colon cancer, nonpolyposis type 1 (E. coli)
- MLH1** (Homo sapiens): mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)

Items 1 - 20 of 941100

Tabs

Taxonomy filter

Gene Information

Top Organisms [Tree]

- Homo sapiens (482635)
- synthetic construct (118957)
- Mus musculus (69474)
- Danio rerio (29495)
- Xenopus laevis (15332)
- Rattus norvegicus (14639)
- Helicobacter pylori (13098)
- Xenopus (Silurana) tropicalis (11721)
- Human immunodeficiency virus 1 (11699)
- uncultured bacterium (9877)
- Helicobacter pylori
- HPKX_438_CAMC1 (7534)
- HPKX_438_AGOC1 (5206)
- Grossmannia clavigena kv1407 (4996)
- Pongo abelii (4643)
- Drosophila melanogaster (1614)
- Felis catus (1176)
- Canis lupus familiaris (1135)
- All other taxa (24999)

More Precise Nucleotides Search

Found 5 nucleotide sequences. Nucleotide [5]

Display Summary Show 20 Sort By Send to

All: 5 Bacteria: 0 RefSeq: 5 mRNA: 5

Items 1 - 5 of 5 One page

Four MLH1 splice variants

1. **Homo sapiens mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli) (MLH1), transcript variant 4, mRNA**
NM_001167618.1 GI:263191732

2. **Homo sapiens mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli) (MLH1), transcript variant 3, mRNA**
NM_001167618.1 GI:263191712

3. **Homo sapiens mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli) (MLH1), transcript variant 2, mRNA**
NM_001167617.1 GI:263191588

4. **Homo sapiens mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli) (MLH1), transcript variant 1, mRNA**
NM_001167617.1 GI:263191588

5. **Homo sapiens mutS homolog 2, colon cancer, nonpolyposis type 1 (E. coli) (MSH2), mRNA**
NM_002051.1 GI:4587760

colon cancer[Title] AND nonpolyposis[Title] AND human[Organism]
AND biomed_mrna[Properties] AND srcdb_refseq[Properties]

Fielded Searching

term1[Field] AND/OR/NOT term2[Field] ...

Field	Databases
[MeSH Terms] – Medical Subject Headings	PubMed
[Organism] – NCBI's Taxonomy	Molecular Databases
[Properties] – molecule type, database source, cellular compartment	Sequence databases
[Filter] – records with links to other databases and more	All databases
[Title] – definition line of sequences, titles of other records	All databases

Queries are automatically mapped to the MeSH and organism vocabularies

Examples

Human RefSeq mRNA sequences with creatine kinase in the title

```
human[organism] AND creatine kinase [Title] AND
srcdb_refseq[Properties] AND biomol_mrna[Properties] AND
creatine kinase[Title]
```

PubMed records about Alzheimer disease genetics published in the past year with free full-text in PubMed Central

```
Alzheimer disease[MeSH Terms] AND genetics[subheading] AND
pubmed_pmc[Filter] AND published last year[Filter]
```

NCBI Field Guide

PubMed: Medical Subject Headings

NCBI MeSH A service of the National Library of Medicine and the National Institutes of Health

Search: MeSH for Alzheimer's Disease

Display: Summary Show 20 Send to

All: 3 Items 1 - 3 of 3 One page.

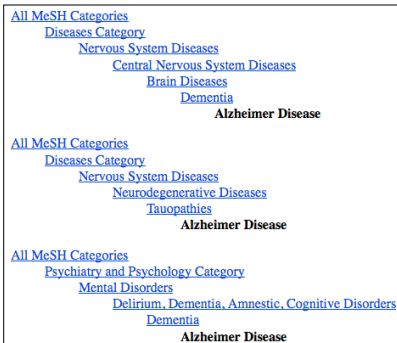
1: Alzheimer Disease [Links](#)
 A degenerative disease of the BRAIN characterized by the insidious onset of DEMENTIA. Impairment of MEMORY, judgment, attention span, and problem solving skills are followed by severe APRAXIAS and a global loss of cognitive abilities. The condition primarily occurs after age 60, and is marked pathologically by severe cortical atrophy and the triad of SENILE PLAQUES, NEUROFIBRILLARY TANGLES, and NEUROFIL THREADS. (From Adams et al., Principles of Neurology, 6th ed, pp1049-57)
 Year introduced: 1998(1963)

2: familial Alzheimer's disease protein 1 [Links](#) **[Substance Name]**
 expressed in lymphoblastoid and brain cells derived from Alzheimer's disease patients; amino acid sequence given in first source; GenBank X81004
 Date introduced: August 4, 1995

3: Alzheimer's disease antigen [Links](#) **[Substance Name]**
 reacts with monoclonal antibody Alz 50; MW 68 kDa
 Date introduced: July 12, 1988

NCBI Field Guide

MeSH is an Ontology



NCBI Field Guide

Organism Field: NCBI's Taxonomy

NCBI Taxonomy

Homo sapiens
 Taxonomy ID: 9606
 Genbank common name: human
 Inherited blast name: primates
 Rank: species
 Genetic code: Translation table 1 (Standard)
 Mitochondrial genetic code: Translation table 2 (Vertebrate Mitochondrial)
 Other names:
 common name: man

Lineage (full)
 cellular organisms; Eukaryotes; Fungi/Metazoa group; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Euarchontoglires; Primates; Haplorhini; Simiiformes; Catarrhini; Hominoidea; Homidae; Homininae; Homo

Entrez records

Database name	Subtree links	Direct links
Nucleotide	5,098,284	5,098,283
Nucleotide EST	8,296,280	8,296,280
Nucleotide GSS	1,232,874	1,231,548
Protein	506,661	506,566
Structure	14,402	14,402
Genome Sequences	15	15
Genome Projects	1	1
Papert	20,379	20,379
SNP	25,003,333	25,000,333
3D Domains	56,372	56,372
Domains	16	16
GEO Datasets	8,084	8,084
GEO Expressions	17,689,684	17,689,684
UniGene	123,356	123,356
UniSTS	327,100	327,100
PubMed Central	6,884	6,882
Gene	45,339	45,302
RefSeqGene	18,876	18,876
SRA Experiments	5,575	5,575
Taxonomy	2	1

Genome information
 See the NCBI Genome homepage
 Go to NCBI genomic BLAST page for Homo sapiens

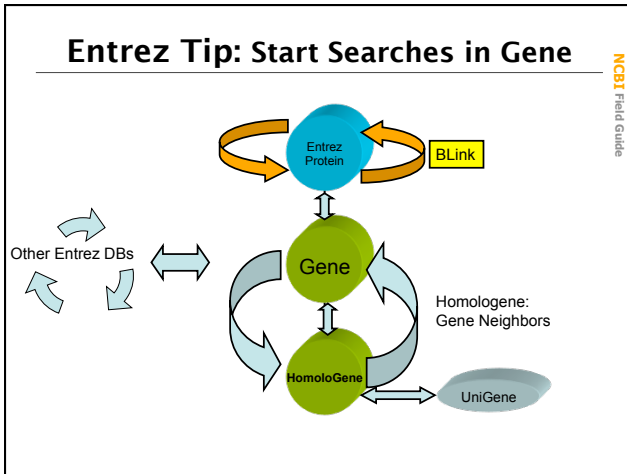
Genome view: 24 chromosomes

Names: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y

See the Mitochondrion Genome
 See the TRACE Assembly

All molecular databases

NCBI Field Guide



Gene Results

Items 1 - 11 of 11

1: **MLH1**
Official Symbol **MLH1** and Name: **mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)** [Homo sapiens]
Other Aliases: **COC1, FOC1, HNPCC, HNPCC1, LCF52**
Other Designations: **mutL homolog 2**
Chromosomes: **2, Location: 3p21.3**
Annotation: **Chromosome 2, NC_000002.11 (4763263, 47716360)**
GeneID: 4292

2: **MLH1**
Official Symbol **MLH1** and Name: **mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)** [Homo sapiens]
Other Aliases: **COC2, FCC2, HNPCC, HNPCC2, MGS172, MLH1**
Other Designations: **DNA mismatch repair protein MLH1, MutL protein homolog 1**
Chromosomes: **3, Location: 3p21.3**
Annotation: **Chromosome 3, NC_000003.11 (37034979..37092337)**
MIM: 120436
GeneID: 4292

3: **MLH1**
Official Symbol **MLH1** and Name: **mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)** [Homo sapiens]
Other Aliases: **COC2, FCC2, HNPCC, HNPCC2, MGS172, MLH1**
Other Designations: **DNA mismatch repair protein MLH1, MutL protein homolog 1**
Chromosomes: **3, Location: 3p21.3**
Annotation: **Chromosome 3, NC_000003.11 (37034979..37092337)**
MIM: 120436
GeneID: 4292

4: **TGFBR2**
Official Symbol **TGFBR2** and Name: **transforming growth factor, beta receptor II (TGFBR2a) (Homo sapiens)**
Other Aliases: **AA73, FAS3, LDS18, LDS28, MFS2, RIIC, TAAO2, TGF-R-2, TGF-beta-RII**
Other Designations: **TGF-beta receptor type IIB, TGF-beta type II receptor, transforming growth factor beta receptor type IIC, transforming growth factor, beta receptor 1**
Chromosomes: **3, Location: 3p22**
Annotation: **Chromosome 3, NC_000003.11 (30847994..30739534)**
MIM: 190182
GeneID: 7048

5: **MLH1**
Official Symbol **MLH1** and Name: **mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)** [Homo sapiens]
Other Aliases: **COC2, FCC2, HNPCC, HNPCC2, MGS172, MLH1**
Other Designations: **DNA mismatch repair protein MLH1, MutL protein homolog 1**
Chromosomes: **3, Location: 3p21.3**
Annotation: **Chromosome 3, NC_000003.11 (37034979..37092337)**
MIM: 120436
GeneID: 4292

6: **MLH1**
Official Symbol **MLH1** and Name: **mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)** [Homo sapiens]
Other Aliases: **COC2, FCC2, HNPCC, HNPCC2, MGS172, MLH1**
Other Designations: **DNA mismatch repair protein MLH1, MutL protein homolog 1**
Chromosomes: **3, Location: 3p21.3**
Annotation: **Chromosome 3, NC_000003.11 (37034979..37092337)**
MIM: 120436
GeneID: 4292

7: **MLH1**
Official Symbol **MLH1** and Name: **mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)** [Homo sapiens]
Other Aliases: **COC2, FCC2, HNPCC, HNPCC2, MGS172, MLH1**
Other Designations: **DNA mismatch repair protein MLH1, MutL protein homolog 1**
Chromosomes: **3, Location: 3p21.3**
Annotation: **Chromosome 3, NC_000003.11 (37034979..37092337)**
MIM: 120436
GeneID: 4292

8: **MLH1**
Official Symbol **MLH1** and Name: **mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)** [Homo sapiens]
Other Aliases: **COC2, FCC2, HNPCC, HNPCC2, MGS172, MLH1**
Other Designations: **DNA mismatch repair protein MLH1, MutL protein homolog 1**
Chromosomes: **3, Location: 3p21.3**
Annotation: **Chromosome 3, NC_000003.11 (37034979..37092337)**
MIM: 120436
GeneID: 4292

9: **MLH1**
Official Symbol **MLH1** and Name: **mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)** [Homo sapiens]
Other Aliases: **COC2, FCC2, HNPCC, HNPCC2, MGS172, MLH1**
Other Designations: **DNA mismatch repair protein MLH1, MutL protein homolog 1**
Chromosomes: **3, Location: 3p21.3**
Annotation: **Chromosome 3, NC_000003.11 (37034979..37092337)**
MIM: 120436
GeneID: 4292

10: **MLH1**
Official Symbol **MLH1** and Name: **mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)** [Homo sapiens]
Other Aliases: **COC2, FCC2, HNPCC, HNPCC2, MGS172, MLH1**
Other Designations: **DNA mismatch repair protein MLH1, MutL protein homolog 1**
Chromosomes: **3, Location: 3p21.3**
Annotation: **Chromosome 3, NC_000003.11 (37034979..37092337)**
MIM: 120436
GeneID: 4292

11: **MLH1**
Official Symbol **MLH1** and Name: **mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)** [Homo sapiens]
Other Aliases: **COC2, FCC2, HNPCC, HNPCC2, MGS172, MLH1**
Other Designations: **DNA mismatch repair protein MLH1, MutL protein homolog 1**
Chromosomes: **3, Location: 3p21.3**
Annotation: **Chromosome 3, NC_000003.11 (37034979..37092337)**
MIM: 120436
GeneID: 4292

nonpolyposis colon cancer AND human[Organism]

Precise Results

Search: **MLH1[Gene Name] AND Human[Organism]**

1: **MLH1 mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli) [Homo sapiens]**

GeneID: 4292 updated 8-Dec-2009

Summary

Official Symbol **MLH1** provided by HGNC

Official Full Name **mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)** provided by HGNC

Primary Source **NC_000002.11**

See related **Ensembl: ENSG0000026242; HPRD:00390; MIM:120436**

Gene type **protein coding**

RefSeq status **REVIEWED**

Organism **Homo sapiens**

Lineage **Eukaryota, Mammalia, Hominoidea, Homo**

Also known as **FCC2; COCA2; HNPCC; MLH1; HNPCC2; MGS172; MLH1**

Summary This gene was identified as a locus frequently mutated in hereditary nonpolyposis colon cancer (HNPCC). It is a human homolog of the E. coli DNA mismatch repair gene mutL, consistent with the characteristic alterations in microsatellite sequences (RER+ phenotype) found in HNPCC. Alternative splicing results in multiple transcript variants encoding distinct isoforms. Additional transcript variants have been described, but their full-length natures have not been determined.

NCBI Field Guide

MLH1 Gene Record

1: **MLH1 mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli) [Homo sapiens]**

GeneID: 4292 updated 8-Dec-2009

Summary

Official Symbol **MLH1** provided by HGNC

Official Full Name **mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)** provided by HGNC

Primary Source **NC_000002.11**

See related **Ensembl: ENSG0000026242; HPRD:00390; MIM:120436**

Gene type **protein coding**

RefSeq status **REVIEWED**

Organism **Homo sapiens**

Lineage **Eukaryota, Mammalia, Hominoidea, Homo**

Also known as **FCC2; COCA2; HNPCC; MLH1; HNPCC2; MGS172; MLH1**

Summary This gene was identified as a locus frequently mutated in hereditary nonpolyposis colon cancer (HNPCC). It is a human homolog of the E. coli DNA mismatch repair gene mutL, consistent with the characteristic alterations in microsatellite sequences (RER+ phenotype) found in HNPCC. Alternative splicing results in multiple transcript variants encoding distinct isoforms. Additional transcript variants have been described, but their full-length natures have not been determined.

GeneRIFs: Gene References Into Function

1. BRCA1 methylation correlated with age at diagnosis (P = .015) and 5-years disease free survival (P = .016) while MLH1 methylation was more frequent in larger tumors (P = .002) and in presence of distant metastasis (P = .004).

2. Co-immunoprecipitation analyses revealed that MutLalpha, and also MSH2 and MSH6, components of the MutLalpha heterodimer, form complexes with Poldelta in human cells.

3. MutLalpha (MLH1-PM2), replication protein A (RPA), and HMG18 have roles in 5'-directed mismatch repair.

4. The present study reveals a possible function of MLH1 protein in protecting colon tumor cells from resistance acquisition by trichostatin A.

5. Assessing high rate microsatellite instability and expressions of MLH1 could be used to distinguish benign and malignant insulinomas and to predict the outcome of patients.

6. New MLH1 missense mutation in Muir-Torre syndrome associated with familial transmission of different gastrointestinal adenocarcinomas.

7. microsatellite instability was associated with MLH1 in gastric carcinomas.

8. microsatellite instability in BAT26 and BAT25 were predictive of MSH1 mutated Lynch syndrome.

9. The association of MLH1 variant forms 1 and 3 did not correlate with the same transcription.

Submit: [New GeneRIF](#) [Correction](#)

NCBI Field Guide

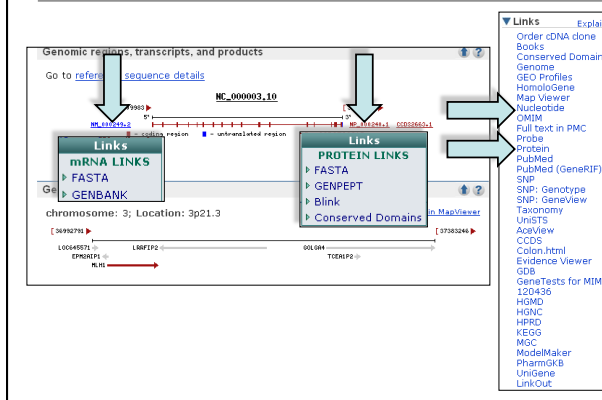
NCBI Field Guide

Interactions					B.3.1		Evidence
Description	Interactant	Other Gene	Complex	Score	Function	Evidence	
Product							
EP21 interacts with the MLH1 promoter.					ATM binding	1EA	
NC_000003.9	NP_055161.1	EP21			RuvA/RuvB complex binding	1EA	
EP24 interacts with the MLH1 promoter region.					puromycin/thiamine repair binding	1DA Published	
NC_000003.9	NP_051941.2	EP24			overactivated DNA binding	1EA	
					protein binding	1PI	
					single-stranded DNA binding	1EA Published	
NP_00240.1	NP_00048.1	BLM					
MLH1 interacts with BLM.					Process	Evidence	
NP_00240.1	NP_00048.1	BLM			MLH1 damages response, signal transduction resulting in induction of apoptosis	1EA	
NP_00240.1	NP_005225.1	BRCA1			cell cycle	1EA	
The endorecise HES1 interacts with the mismatch repair protein MLH1.					double-strand break repair via nonhomologous end joining	1EA	
NP_00240.1	NP_036773.7	MLH1			dephos phosphorylation	1EA	
The endorecise HESD10 interacts with the mismatch repair protein MLH1.					mark meiosis chromosome segregation	1EA	
NP_00240.1	NP_05013.3	ESD1			meiotic metabolism, 1 phase progression	1EA	
					mismatch repair	1EA	
					negative regulation of mitotic recombination	1EA	
					nuclear transcription inhibitory activity DNA shunting	1EA	
NP_00240.1	NP_56908.1	XPO1			negative regulation of mitotic recombination	1EA	
					positive regulation of mitotic recombination	1EA	
NP_00240.1	NP_00316.1	MDSB4			response to DNA damage stimulus	1EA	
MLH1 and interacts with MDSB4.					actin cytoskeleton	1EA	
NP_00240.1	NP_032615.1	MLH3			regulation of meiotic recombination intermediates	1EA	
					homologous recombination	1EA	
NP_00240.1	BA082353.1	MDSB3			homologous recombination	1EA	
					epistatic hypermutation of immunoglobulin genes	1EA	
					homologous recombination	1EA	
					spindle midzone assembly involved in meiosis	1EA	
					Component	Evidence	
					RuvA/RuvB complex	1EA	
					chromatin	1EA	
					male germ cell nucleus	1EA	
					nucleus	1C Published	
					overactivated complex	1EA	

NCBI Field Guide

Genomes			RefSeq of Annotated Genomes: Build 37.1					
1. NC_007039.1 RefSeqGenome	Range	84567-82,009	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_007039.1						
NC_007039.1 RefSeqGenome	Range	84567-82,009	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_007039.1						
Genome Reference Consortium Human Build 37 (GRCh37), Primary Assembly								
Genomes and Proteins								
1. NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000	Download	GenBank	RefSeq			
Downloaded	GenBank	NC_009003.1						
NC_009003.1 RefSeqGenome	Range	1,000,000-1,000,000						

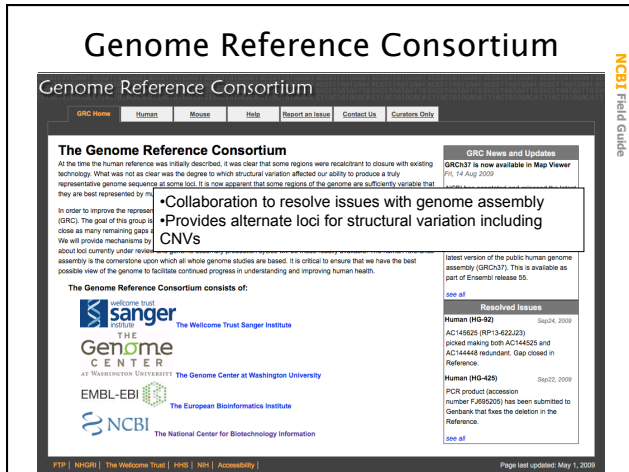
NCBI Field Guide



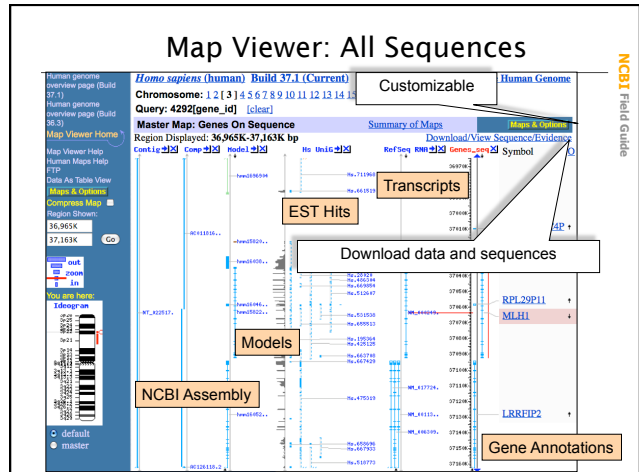
NCBI Field Guide

[illegible]

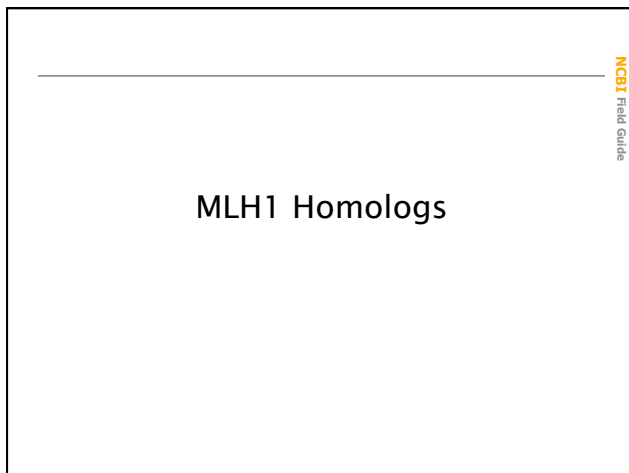
NCBI Field Guide



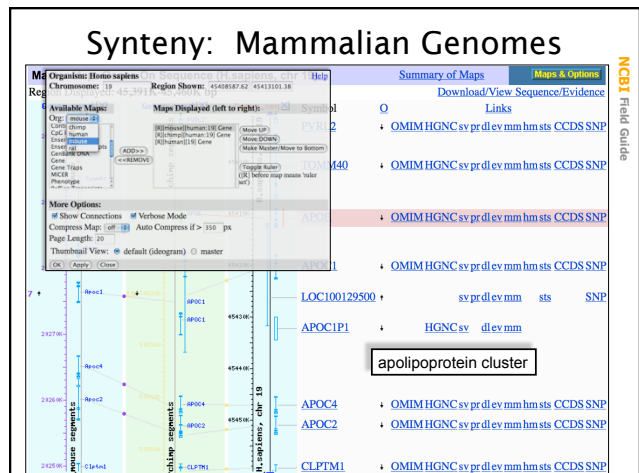
NCBI Field Guide



NCBI Field Guide



NCBI Field Guide



Finding Homologs: HomoloGene

Links

- Order cDNA clone
- Books
- Conserved Domains
- Genome
- GEO Profiles
- HomoloGene
- Map Viewer
- Nucleotide
- OMIM
- Full text in PMC
- Probe
- Protein
- PubMed
- PubMed (GeneRIF)
- SNP
- SNP Genotype
- SNP GeneView
- Taxonomy
- UniSTS
- AcycView
- CCDS
- Coln.html
- Evidence Viewer
- GD
- Gene Tests for MIM
- LOC55
- HGM
- HNC
- HPD
- KEGG
- MOC
- ModelMaker
- PharmWiki
- UniProt
- LinkOut

Protein record

More about the MLH1 gene

This gene was identified as a locus frequently mutated in hereditary nonpolyposis colon cancer (HNPCC). It is a human homolog of the E. coli...

Homologs of the MLH1 gene

The MLH1 gene is conserved in chimpanzee, dog, cow, mouse, rat, chicken, zebrafish, fruit fly, mosquito, Drosophila, M. grisea, N. crassa, and P. falciparum.

Homologene Ad

NCBI Field Guide

Discovery column

HomoloGene Cluster

Links

- BioSystems
- Conserved Domains
- Genome
- GEO Profiles
- Map Viewer
- Nucleotide
- PubChem Compound
- PubChem Substance
- Full text in PMC
- Probe
- Protein
- PubMed
- PubMed (GeneRIF)
- Taxonomy
- UniGene
- FlyBase

Links

- Related Sequences
- Domain Relatives
- Identical Proteins
- BioSystems
- Conserved Domains
- Gene
- Genome Project
- Map Viewer
- Nucleotide
- Full text in PMC
- PubMed
- PubMed (RefSeq)
- PubMed (Weighted)
- Related Structure
- Taxonomy
- UniGene
- BLINK

Protein Links

HomoloGene Downloader

Download **Protein** **mRNA** **Genomic**

Select which sequences should be included:

(Select All) (Unselect All)

Species	Gene	Accession
<input checked="" type="checkbox"/> H.sapiens	MLH1	NM_002429.2
<input checked="" type="checkbox"/> P.troglodytes	MLH1	XM_001170433.1
<input checked="" type="checkbox"/> C.lupus	MLH1	NM_534219.2
<input checked="" type="checkbox"/> B.taurus	MLH1	NM_001705052.1
<input checked="" type="checkbox"/> M.musculus	Mh1	NM_026910.1
<input checked="" type="checkbox"/> R.norvegicus	Mh1	NM_031053.1
<input checked="" type="checkbox"/> G.gallus	MLH1	NM_418228.1
<input checked="" type="checkbox"/> D.rerio	mh1	NM_200559.1
<input checked="" type="checkbox"/> D.melanogaster	Mh1	NM_057842.2
<input checked="" type="checkbox"/> A.gamabae	AgpP_AGAP012192	NM_320342.4
<input checked="" type="checkbox"/> S.pombe	mh1	NM_01022118.1
<input checked="" type="checkbox"/> S.cerevisiae	MLH1	MLM1_6323819
<input checked="" type="checkbox"/> Klacids	KLA000995g	NM_433041.1
<input checked="" type="checkbox"/> E.gosypii	AGOS_AFL199C	NM_210705.1
<input checked="" type="checkbox"/> M.grisea	MG0_07121	NM_367194.1
<input checked="" type="checkbox"/> N.crassa	NCU0309.1	NM_367194.1
<input checked="" type="checkbox"/> P.falciparum	PF11_0184	NM_001347819.1

Protein **mRNA** **Genomic**

hg1 [4557757]ref[HP_000240.1] Mult. protein homolog 1 isoform 1 [Homo sapiens]

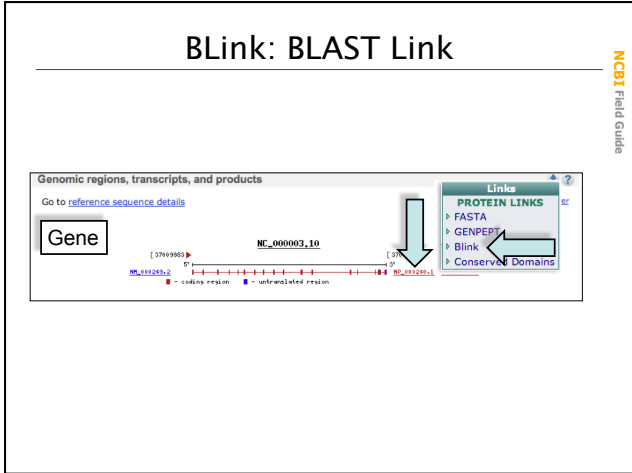
MLH1 protein record

MLH1 protein record

Finding Protein Homologs

NCBI Field Guide

NCBI Field Guide



NCBI Field Guide

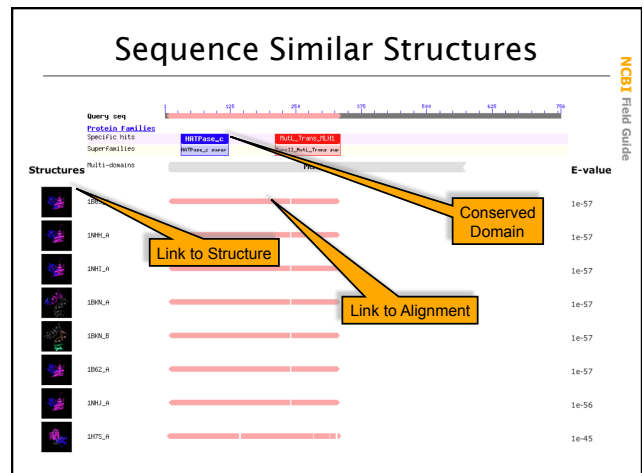
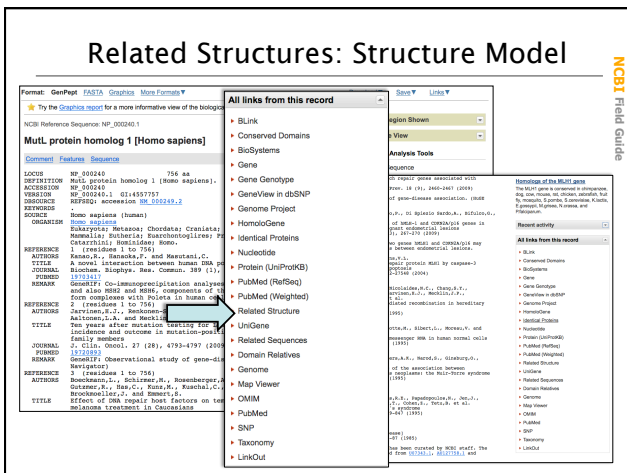
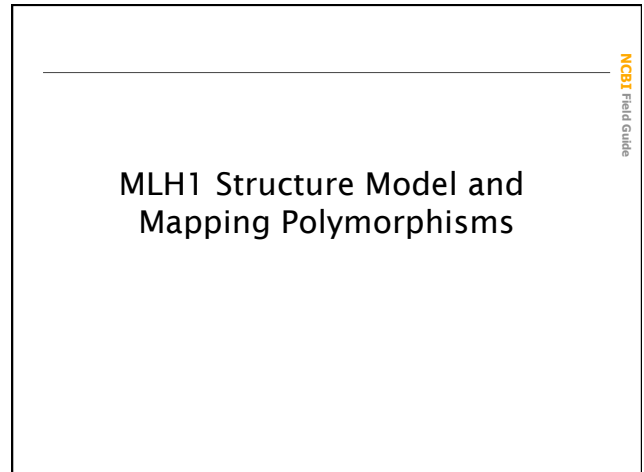
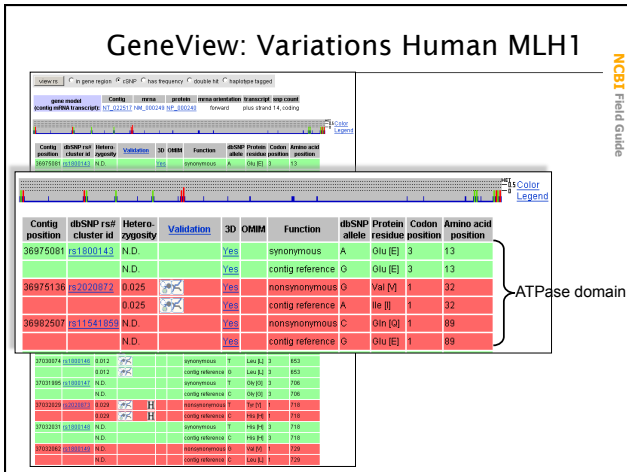
[illegible]

9205

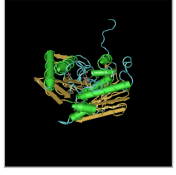
[illegible]

NCBI Field Guide

The screenshot shows the NCBI Fold Guide interface for the protein 7E6. The 'Gene Links' section is highlighted with a blue arrow pointing to the 'Gene' link, and the 'Protein Links' section is highlighted with a blue arrow pointing to the 'Protein' link. The interface includes a top navigation bar with 'GenPept', 'FASTA', 'Graphics', and 'More Formats'. The 'Gene Links' section lists various databases like PubMed, Gene, and UniProt. The 'Protein Links' section lists various databases like UniProt, PDB, and RefSeq.



E. coli MutL Structure



MMDB ID: 10447 **PDB ID:** 1B63

Reference: Ban C, Junop M, Yang W. Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair. Cell. 1997; 89: 55-67.

Description: MutL Complexed With Adpnp.

Taxonomy: Escherichia coli

Related Structures: VAST

Tasks: [Download] [Drawings] [All Atoms] [View Cn3D Tutorial]

Molecular components in the MMDB structure are listed below and may include macromolecular chains, 3D domains, protein classifications (domain families and superfamily) and specific hits. Mouse over each icon for more information on the component.

Protein: MutL
Domain families: MutL-like
Superfamilies: MutL-like

Cn3D viewer

Conserved Domain

MutL protein homolog 1 [Homo sapiens]

Related Structures:

Alignment Based Model: Mapping Polymorphisms

Query sequence: [gi45577577:NM_000240.1](#)
MutL protein homolog 1 [Homo sapiens]

Related structure: [1B63_A](#) [Search references in PubMed]
Chain A: MutL Complexed With Adpnp

Alignment to query sequence

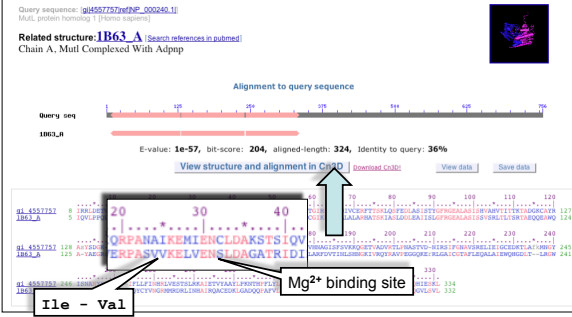
Query seq: 1 100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000 2100 2200 2300 2400 2500 2600 2700 2800 2900 3000 3100 3200 3300 3400 3500 3600 3700 3800 3900 4000 4100 4200 4300 4400 4500 4600 4700 4800 4900 5000 5100 5200 5300 5400 5500 5600 5700 5800 5900 6000 6100 6200 6300 6400 6500 6600 6700 6800 6900 7000 7100 7200 7300 7400 7500 7600 7700 7800 7900 8000 8100 8200 8300 8400 8500 8600 8700 8800 8900 9000 9100 9200 9300 9400 9500 9600 9700 9800 9900 10000

E-value: 1e-57, bit-score: 204, aligned-length: 324, Identity to query: 36%

View structure and alignment in Cn3D [Download Cn3D] [View data] [Save data]

Ile - Val

Mg²⁺ binding site



Better Model: Conserved Domain

Genomic regions, transcripts, and products

(plus) Go to [reference sequence details](#)

Gene

MutL protein homolog 1 [Homo sapiens]

Protein

Related Structures

Links

PROTEIN LINKS

FASTA

GENPEPT

Blink

Conserved Domains

Customize View

Sequence Analysis Tools

BLAST

Conserved Domains

Articles about the MutL gene

Cancer risk in a cohort of subjects carrying a single m [J Clin Oncol. 2005]

Better Model: Conserved Domain

cd00075.3

Links:

Source: Smart

Taxonomy: root

PubMed: 2 links

Protein: cd00075 related architectures

Representatives:

Related CD: 3 links

Statistics:

PSSM-16: 28956

View PSSM: cd00075

Aligned: 483 rows

PSSM: 103 columns

Status: curated CD

Created: 1-Nov-2000

Updated: 11-Apr-2003

Structure:

Show Structure

Draw: [All atoms] [up to 10]

Aligned: [up to 10]

Download Cn3D

ATP binding **Mg²⁺ binding** **G-X-G motif**

Feature 1: ATP binding site

Evidence:

- Structure: 1A4H bound Geldanamycin (ATP analog)
- Comment: the ATP-lid is open leaving and the bound ATP completely exposed to the solvent.
- Citation: PMID 10637609
- Citation: PMID 9230303
- Structure: 1B7Q, A bound ADP-Mg²⁺
- View structure with Cn3D

Feature 1

Ile - Val

Position 32

Mg²⁺ binding site